Strathmore
UNIVERSITY

INSTITUTE OF MATHEMATICAL SCIENCES (IMS)
MASTER OF SCIENCE IN BIOMATHEMATICS
END OF SEMESTER EXAMINATION
BMA 8408: HEALTH INFORMATICS AND DATA ANALYTICS

DATE: 13th August 2018 Time: 2 Hours

Instructions

1. This examination consists of **FIVE** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question One

1.1 Explain the differences between three major sub-disciplines of Biomedical Informatics namely; bio-informatics and health informatics (4 marks).

1.2 For a classification problem, the aim is to find the best classifier function *h(.)* for a data set *D* using a loss function. State and explain three loss functions (6 marks).

1.3 The following are purchase transactions of different items;

{Blue band, Bread}; {Blue band, Butter, Tea Bags, Sugar}, {Blue band, Bread, Butter, Milk}, {Bread, Butter},
{Bread, Milk, Sugar}, {Blue band, Butter, Milk, Sugar}, {Blue band, Tea Bags, Sugar}, {Blue band, Bread, Butter},
{Bread, Milk, Cheese}, {Blue band, Bread, Butter, Cheese}

Calculate the support, confidence and lift of the following association rule *{Blue band} -> {Tea Bags, Sugar}. (6 marks)*

1.4 K-Nearest Neighbor selects the nearest neighbor as a prediction if K=1. Explain how KNN approach makes prediction for K>1 for both classification and regression problems. (2 marks)

1.5 Major practical application of Random Forest and Association Rule apart from prediction (2 marks)

Question Two

2.1 Explain the meaning of overfitting as applied in prediction. (2 marks)

2.2 Describe two types of decision trees and explain different nodes of a decision tree (10 marks)

2.3 Explain the term "random" as used in Random Forest predictions (2 marks)

2.4 In random forest prediction, we do not need to create "train" and "test" data sets in developing model. Explain (4 marks)

2.5 Name two important parameters used in fine tuning random forest (2marks)

Question Three

3.1 Explain why a confidence of 99% in association rules is more interesting than a confidence of 100%. (6 marks)

3.2 The algorithm that we used to do association rule mining is the apriori algorithm. This algorithm is efficient because it relies on and exploits the apriori property. What is the apriori property? (2 marks)

3.3 Briefly describe the general objective of Association Rules mining. (2 marks)

3.4 Explain different layers in artificial neural network (ANN) (6 marks)

3.5 Explain the meaning of k-fold Cross-Validation (4 marks)

Question Four

4.1 What are support vectors? (2 marks)

4.2 Explain two parameters (gamma and cost) when using radial kernel in support vector machines (SVM) with high Gamma value. What does this signify? (4 marks)

4.3 If you achieve 100% accuracy on my training set, and 70% on validation set, what should you look out for? (2 marks)

4.4 What are two real world applications of the SVM? (2 marks)

4.5 Explain the difference between supervised, unsupervised and semi-supervised learning. (6 marks)

4.6 Explain Hard Margin and the short coming of its use in support vector machine (2 marks)